

LISTING OF THE CLAIMS

Claims 1-30 are pending. No claims are amended. No claims are cancelled or added. The following listing of claims replaces all prior versions and listings of claims in the application.

1. (Original) A method for clustering queries, the method comprising:
identifying a same document and/or a plurality of similar documents
selected by a user in response to a plurality of queries; and
responsive to identifying the same document and/or the similar documents,
generating a query cluster to indicate that the queries are similar independent of
whether individual ones of the queries comprise similar composition with respect
to other ones of the queries.

2. (Original) A method as recited in claim 1, wherein the queries
comprise a well formed natural language question, a keyword, or a phrase.

3. (Original) A method as recited in claim 1, wherein the query cluster
is used to disambiguate a word or phrase in a query of the queries.

4. (Original) A method as recited in claim 1, further comprising
determining that the queries are similar based on similar keyword or phrase
composition.

1 5. (Original) A method as recited in claim 1, wherein identifying the
2 same document and/or the similar documents further comprises:

3 determining the similar documents by evaluating a set of selected similar
4 documents chosen responsive to queries p and q of the queries, wherein
5 documents $D_C(.)$ is a subset of a result list $D(.)$ according to the following:

6
$$D_C(p) = \{ d_{p1}, d_{p2}, \dots, d_{pi} \} \subseteq D(p)$$

7
$$D_C(q) = \{ d_{q1}, d_{q2}, \dots, d_{qj} \} \subseteq D(q);$$

8 wherein similarity based on selection of documents is based on:

9 If $D_C(p) \cap D_C(q) = \{ d_{pq1}, d_{pq2}, \dots, d_{pqk} \} \neq \emptyset$, then documents $d_{pq1},$
10 d_{pq2}, \dots, d_{pqk} represent a set of common topics of queries p and q , and,

11 whereby the similar documents between queries p and q is determined by
12 $D_C(p) \cap D_C(q).$

13
14 6. (Original) A method as recited in claim 1, further comprising
15 constructing a thesaurus comprising a plurality of synsets, wherein each synset
16 comprises one or more query clusters.

17
18 7. (Original) A method as recited in claim 1, wherein identifying the
19 same document and/or the similar documents further comprises determining the
20 similar documents based on a proportionality of commonly selected individual
21 documents.
22
23
24
25

1 8. (Original) A method as recited in claim 7, wherein identifying the
2 same document and/or the similar documents further comprises:

3 determining the similar documents based on a proportionality of commonly
4 selected individual documents, such that:

5
$$\text{similarity}_{\text{single_doc}}(p, q) = \frac{RD(p, q)}{\text{Max}(rd(p), rd(q))},$$

6 wherein $rd(.)$ is the number of clicked documents for a query of the queries,
7 and wherein $RD(p, q)$ is the number of document selections in common.
8

9 9. (Original) A method as recited in claim 1, wherein identifying the
10 same document and/or the similar documents further comprises:

11 determining the similar documents based on a hierarchical positioning
12 between individual ones of a plurality of documents commonly selected across the
13 queries.
14
15
16
17
18
19
20
21
22
23
24
25

10. (Original) A method as recited in claim 9:

wherein $F(d_i, d_j)$ is a lowest common parent node for documents d_i and d_j ;

wherein $L(x)$ is a level of a node x ;

wherein L_Total identifies a total number of levels in a hierarchy; and

wherein a similarity between two documents is defined as follows:

$$s(d_i, d_j) = \frac{L(F(d_i, d_j)) - 1}{L_Total - 1}, \text{ such that}$$

$$s(d_i, d_j) = 1; \text{ and } s(d_i, d_j) = 0 \text{ if } F(d_i, d_j) = \text{root}; \text{ and}$$

the method further comprises:

incorporating $s(d_i, d_j)$ into a calculation of query similarity, wherein.

d_i ($1 \leq i \leq m$) and d_j ($1 \leq j \leq n$) be a set of selected documents for queries p and q respectively such that:

$$similarity_{hierarchy}(p, q) = \frac{1}{2} \times \left(\frac{\sum_{i=1}^m (\max_{j=1}^n s(d_i, d_j))}{rd(p)} + \frac{\sum_{j=1}^n (\max_{i=1}^m s(d_i, d_j))}{rd(q)} \right)$$

11. (Original) Computer-readable media comprising computer-executable instructions for identifying similar queries, the computer-executable instructions comprising instructions for:

identifying a same document and/or a plurality of similar documents selected by a user in response to a plurality of queries; and

responsive to identifying the same document and/or the similar documents, generating a query cluster to indicate that the queries are similar independent of

1 whether individual ones of the queries comprise similar composition with respect
2 to other ones of the queries.

3
4 12. (Original) Computer-readable media as recited in claim 11, wherein
5 the queries comprise a well formed natural language question, a keyword, or a
6 phrase.

7
8 13. (Original) Computer-readable media as recited in claim 11, wherein
9 the query cluster is used to disambiguate a word or phrase in a query of the
10 queries.

11
12 14. (Original) Computer-readable media as recited in claim 11, wherein
13 the computer-executable instructions further comprise instructions for determining
14 that the queries are similar based on similar keyword or phrase composition.

15. (Original) Computer-readable media as recited in claim 11, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for:

determining the similar documents by evaluating a set of selected similar documents chosen responsive to queries p and q of the queries, wherein documents $D_C(.)$ is a subset of a result list $D(.)$ according to the following:

$$D_C(p) = \{ d_{p1}, d_{p2}, \dots, d_{pi} \} \subseteq D(p)$$

$$D_C(q) = \{ d_{q1}, d_{q2}, \dots, d_{qj} \} \subseteq D(q);$$

wherein similarity based on selection of documents is based on:

If $D_C(p) \cap D_C(q) = \{ d_{pq1}, d_{pq2}, \dots, d_{pqk} \} \neq \emptyset$, then documents $d_{pq1}, d_{pq2}, \dots, d_{pqk}$ represent a set of common topics of queries p and q , and,

whereby the similar documents between queries p and q is determined by $D_C(p) \cap D_C(q)$.

16. (Original) Computer-readable media as recited in claim 11, wherein the computer-executable instructions further comprise instructions for constructing a thesaurus comprising a plurality of synsets, wherein each synset comprises one or more query clusters.

17. (Original) Computer-readable media as recited in claim 11, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for determining the similar documents based on a proportionality of commonly selected individual documents.

18. (Original) Computer-readable media as recited in claim 17, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for:

determining the similar documents based on a proportionality of commonly selected individual documents, such that:

$$\text{similarity}_{\text{single_doc}}(p, q) = \frac{RD(p, q)}{\text{Max}(rd(p), rd(q))},$$

wherein $rd(.)$ is the number of clicked documents for a query of the queries, and wherein $RD(p, q)$ is the number of document selections in common.

19. (Original) Computer-readable media as recited in claim 11, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for:

determining the similar documents based on a hierarchical positioning between individual ones of a plurality of documents commonly selected across the queries.

20. (Original) Computer-readable media as recited in claim 19:

wherein $F(d_i, d_j)$ is a lowest common parent node for documents d_i and d_j ;

wherein $L(x)$ is a level of a node x ;

wherein L_Total identifies a total number of levels in a hierarchy; and

wherein a similarity between two documents is defined as follows:

$$s(d_i, d_j) = \frac{L(F(d_i, d_j)) - 1}{L_Total - 1}, \text{ such that}$$

$$s(d_i, d_i) = 1; \text{ and } s(d_i, d_j) = 0 \text{ if } F(d_i, d_j) = \text{root}; \text{ and}$$

wherein the computer-executable instructions further comprise instructions for:

incorporating $s(d_i, d_j)$ into a calculation of query similarity, wherein d_i ($1 \leq i \leq m$) and d_j ($1 \leq j \leq n$) be a set of selected documents for queries p and q respectively such that:

$$\text{similarity}_{\text{hierarchy}}(p, q) = \frac{1}{2} \times \left(\frac{\sum_{i=1}^m (\max_{j=1}^n s(d_i, d_j))}{rd(p)} + \frac{\sum_{j=1}^n (\max_{i=1}^m s(d_i, d_j))}{rd(q)} \right)$$

1 **21. (Original) A computing device comprising:**

2 a processor coupled to a memory, the memory comprising computer
3 executable instructions, the processor being configured to fetch and execute
4 the computer-executable instructions for:

5 identifying a same document and/or a plurality of similar documents
6 selected by a user in response to a plurality of queries; and

7 responsive to identifying the same document and/or the similar
8 documents, generating a query cluster to indicate that the queries are similar
9 independent of whether individual ones of the queries comprise similar
10 composition with respect to other ones of the queries.

11
12 **22. (Original) A computing device as recited in claim 21, wherein the**
13 **queries comprise a well formed natural language question, a keyword, or a phrase.**

14
15 **23. (Original) A computing device as recited in claim 21, wherein the**
16 **query cluster is used to disambiguate a word or phrase in a query of the queries.**

17
18 **24. (Original) A computing device as recited in claim 21, wherein the**
19 **computer-executable instructions further comprise instructions for determining**
20 **that the queries are similar based on similar keyword or phrase composition.**
21
22
23
24
25

25. (Original) A computing device as recited in claim 21, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for:

determining the similar documents by evaluating a set of selected similar documents chosen responsive to queries p and q of the queries, wherein documents $D_C(.)$ is a subset of a result list $D(.)$ according to the following:

$$D_C(p) = \{d_{p1}, d_{p2}, \dots, d_{pi}\} \subseteq D(p)$$

$$D_C(q) = \{d_{q1}, d_{q2}, \dots, d_{qj}\} \subseteq D(q);$$

wherein similarity based on selection of documents is based on:

If $D_C(p) \cap D_C(q) = \{d_{pq1}, d_{pq2}, \dots, d_{pqk}\} \neq \emptyset$, then documents $d_{pq1}, d_{pq2}, \dots, d_{pqk}$ represent a set of common topics of queries p and q , and,

whereby the similar documents between queries p and q is determined by $D_C(p) \cap D_C(q)$.

26. (Original) A computing device as recited in claim 21, wherein the computer-executable instructions further comprise instructions for constructing a thesaurus comprising a plurality of synsets, wherein each synset comprises one or more query clusters.

27. (Original) A computing device as recited in claim 21, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for determining the similar documents based on a proportionality of commonly selected individual documents.

28. (Original) A computing device as recited in claim 27, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for:

determining the similar documents based on a proportionality of commonly selected individual documents, such that:

$$\text{similarity}_{\text{single_doc}}(p, q) = \frac{RD(p, q)}{\text{Max}(rd(p), rd(q))},$$

wherein $rd(\cdot)$ is the number of clicked documents for a query of the queries, and wherein $RD(p, q)$ is the number of document selections in common.

29. (Original) A computing device as recited in claim 21, wherein the instructions for identifying the same document and/or the similar documents further comprise instructions for:

determining the similar documents based on a hierarchical positioning between individual ones of a plurality of documents commonly selected across the queries.

30. (Original) A computing device as recited in claim 29:

wherein $F(d_i, d_j)$ is a lowest common parent node for documents d_i and d_j ;

wherein $L(x)$ is a level of a node x ;

wherein L_Total identifies a total number of levels in a hierarchy; and

wherein a similarity between two documents is defined as follows:

$$s(d_i, d_j) = \frac{L(F(d_i, d_j)) - 1}{L_Total - 1}, \text{ such that}$$

$$s(d_i, d_i) = 1; \text{ and } s(d_i, d_j) = 0 \text{ if } F(d_i, d_j) = \text{root}; \text{ and}$$

wherein the computer-executable instructions further comprise instructions for:

incorporating $s(d_i, d_j)$ into a calculation of query similarity, wherein.

d_i ($1 \leq i \leq m$) and d_j ($1 \leq j \leq n$) be a set of selected documents for queries p and q respectively such that:

$$similarity_{hierarchy}(p, q) = \frac{1}{2} \times \left(\frac{\sum_{i=1}^m (\max_{j=1}^n s(d_i, d_j))}{rd(p)} + \frac{\sum_{j=1}^n (\max_{i=1}^m s(d_i, d_j))}{rd(q)} \right)$$